RESEARCH LETTER

# Genome sequence of *Streptococcus mutans* bacteriophage M102

Jan R. van der Ploeg

Institute of Oral Biology, University of Zürich, Zürich, Switzerland

**Correspondence:** Jan R. van der Ploeg, Institute of Oral Biology, University of Zürich, Plattenstrasse 11, CH-8032, Zürich, Switzerland. Tel.: +41 44 634 3329; fax: +41 44 634 4310; e-mail: jan.vanderploeg@zzmk.uzh.ch

Received 4 May 2007; revised 3 July 2007; accepted 4 July 2007.
First published online August 2007.

DOI:10.1111/j.1574-6968.2007.00873.x

Editor: Oscar Kuipers

**Keywords**
caries; *Streptococcus mutans*; lysin; bacteriophage; holin.

## Abstract

Bacteriophage M102 is a lytic phage specific for serotype c strains of *Streptococcus mutans*, a causative agent of dental caries. In this study, the complete genome sequence of M102 was determined. The genome is 31 147 bp in size and contains 41 ORFs. Most of the ORFs encoding putative phage structural proteins show similarity to those from bacteriophages from *Streptococcus thermophilus*. Bioinformatic analysis indicated that the M102 genome contains an unusual lysis cassette, which encodes a holin and two lytic enzymes.

## Introduction

Dental caries is a frequent disease caused by microorganisms present on the tooth surface that convert carbohydrates present in the diet to lactic acid. The acid production results in demineralization of the tooth enamel and tooth dentin. *Streptococcus mutans*, a gram-positive facultative anaerobic bacterium, has a major role in this demineralization process (Loesche, 1986).

Prevention of dental caries by reduction of *S. mutans* from dental plaque involves nonspecific approaches such as tooth brushing and the use of antimicrobial agents like chlorhexidine. Specific approaches include active or passive immunization and replacement of *S. mutans* by non-cariogenic competitor organisms (Samaranayake, 2002), although none of these concepts has been realized thus far. In theory, bacteriophages or, preferably, lytic enzymes from bacteriophages might be used to specifically remove *S. mutans* from dental plaque as well. The application of bacteriophage and bacteriophage-encoded lytic enzymes as antibacterial agents has recently regained interest (Fischetti, 2001, 2005; Loeffler *et al.*, 2001; Loessner, 2005).

Relatively limited information is available with regard to the ecological role of bacteriophages in the oral cavity. The isolation of bacteriophages from human saliva or dental plaque has had variable success (Tylenda *et al.*, 1985; Armau *et al.*, 1988; Bachrach *et al.*, 2003; Hitch *et al.*, 2004). Lysis of and virus release by several strains of *S. mutans* after mitomycin treatment or UV exposure has been observed, indicating that these strains contained prophages (Greer *et al.*, 1971; Klein & Frank, 1973; Higuchi *et al.*, 1982). However, none of these phages have been purified and further characterized. By screening more than 1000 plaque samples for lytic activity against strains of *S. mutans* and *Streptococcus sobrinus*, Armau *et al.* (1988) isolated 16 lytic bacteriophages. Three of these, phages M102, e1 and f2 were found to be specific for *S. mutans* of serotype c, e and f, respectively (Delisle & Rostkowski, 1993). In the present report, the sequencing and analysis of the genome of phage M102 is described. In addition, the sequence of the lysis genes of phage M101, which is similar to phage M102, was determined.

## Materials and methods

### Bacterial strains, bacteriophages and growth conditions

Bacteriophages M101 and M102 and their host strain *S. mutans* OMZ381 were acquired from G. Tiraby

(Université Paul Sabatier, Toulouse, France). For phage propagation, *S. mutans* OMZ381 was grown in M1D medium, which consisted of 10 g Bacto tryptone, 5 g Bactopeptone, 5 g yeast extract, 5 g NaCl, 2.5 g MOPS (4-morpholinepropanesulfonic acid) and 2 g glucose $L^{-1}$. The pH of M1D was adjusted to pH of 7.4 with NaOH. Solid media contained 1.5% agar (for plates) or 0.7% agar (for soft agar). Phage lysates of M101 and M102 were obtained by infection of exponentially growing cultures of *S. mutans* OMZ381 and subsequent incubation for 16 h at 37 °C. To remove debris, lysates were centrifuged for 10 min at 7000 ***g*** and passed through a 0.4 μm filter.

### Electron microscopy

A drop of phage solution (about $10^9$ PFU $mL^{-1}$) was applied for 3 min to 400 mesh copper grids coated with Formvar (Electron Microscopy Sciences) and stabilized with carbon. The grids were air-dried and subsequently a drop of 1% phosphotungstic acid (PTA) pH 4.4 was applied for 45 s. The PTA solution was removed with filter paper, the grids air dried, and examined with a Philips EM 400T TEM at 80 kV.

### Isolation of phage DNA, DNA manipulations and sequencing

Phage DNA was isolated using the Lambdaprep kit (Promega, Wallisellen, CH). A shotgun library of phage M102 genomic DNA was prepared by GATC Biotech (Konstanz, Germany) as follows. Genomic DNA from phage M102 was sheared by nebulization, blunted with T4 DNA polymerase and Klenow polymerase, and then cloned into pCR4Blunt-Topo (Invitrogen). Plasmids from resulting clones were isolated and sequenced with the T7 and T3 primer. Sequencing was performed using dye terminator technology on a model 3100 sequencer (Applied Biosystems). Sequences were assembled with the program Seqman of the Lasergene package (GATC Biotech). Gaps were closed using PCR and by direct sequencing of the resulting products. A total of 320 sequencing reactions were performed to obtain the complete sequence of M102 with an average coverage of 8.15. Digestion of phage DNA with several different restriction enzymes (PstI, SalI, XhoI, HindIII, EcoRV) yielded fragments whose sizes were in accordance with the genome sequence.

Partial sequences of phage M101 were obtained by direct sequencing of PCR products, obtained with primers derived from the M102 sequence.

### Sequence analysis

The assembled sequence of M102 was analyzed for the presence of ORFs using the program GENEMARK.HMM (Lukashin & Borodovsky, 1998), which uses a cutoff of 42 nucleotides for the minimal coding region. Further sequence analysis used the programs from the GCG package (Accelrys, Cambridge, UK). Blast sequence similarity searches were carried out at http://nbc3.biologie.uni-kl.de/. For multiple sequence alignments, CLUSTALW was used (http://www.ebi.ac.uk/clustalw/).

Phylogenetic distances between phage proteomes were essentially calculated as described (Rohwer & Edwards, 2002). In brief, ORFs from M102 were compared pariwise with all the ORFs of 17 selected *Streptococcal* or *Lactococcal* phages by BLAST analysis (cutoff 0.1). Similar proteins were then aligned using CLUSTALW (gap opening penalty of 10.00 and gap extension penalty of 0.2). The output (in PHYLIP format) was used to determine the phylogenetic distance between each ORF from phage M102 and the corresponding ORF from the other phages with the program PROTDIST (http://artedi.ebc.uu.se/programs/protdist.html). In case of no matching ORF, a penalty of 10 was used. For each phage compared with M102, the sum of the PROTDIST values divided by the number of ORFs used in the comparison was calculated. Calculations were carried out independently for the complete M102 proteome, for Orf1 to Orf18 (structural module) and for Orf19 to Orf41 (lysis and replication module).

## Results and discussion

### Morphology of bacteriophage M102

Electron microscopic analysis (Fig. 1) showed that M102 had a tail with a length of $269 \pm 33$ nm ($n = 69$) and a width of $9.5 \pm 1.2$ nm ($n = 32$). The uniformity of tail lengths indicates that it is noncontractile. The icosahedral phage head had a diameter of $63 \pm 3$ nm ($n = 50$). These values differ somewhat from those reported previously (Delisle & Rostkowski, 1993). The tail was segmented and consisted of

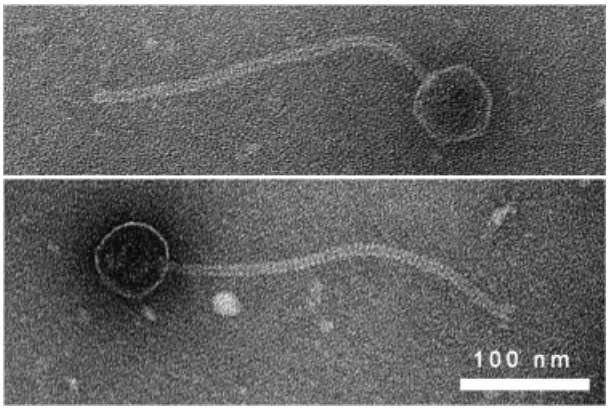


**Fig. 1.** Electron microscopy images of phage M102.

segments of $4.0 \pm 0.2$ nm ($n = 56$). These results indicate that phage M102 belongs to the family of *Siphoviridae* with morphotype B1.

## Genome sequence of M102

The genome of bacteriophage M102 was 31 147 bp in size, which is slightly smaller than the size that was determined previously by restriction enzyme analysis (Delisle & Rostkowski, 1993). The GC content was 39.21%, close to the previously reported value (Delisle & Rostkowski, 1993), but somewhat higher than the value of 36.82% for the genome of *S. mutans* UA159 (Ajdic *et al.*, 2002). Analysis of the genome sequence revealed the presence of 41 ORFs, all transcribed in the same direction (Fig. 2). Most of the ORFs had an ATG startcodon, but there were three ORFs with a GTG start codon and three with a TTG start codon (Table 1). The sequences of the ORFs were compared with sequences from protein databases using protein–protein BLAST. Based on these comparisons, most of the ORFs could be assigned to the different functional groups (Fig. 2). The ORFs from the same functional groups clustered together on the genome.

## Structural module of M102

The first ORF located downstream of the M102 *cos* site, a probable HNH endonuclease, is similar to Orf45 from *Streptococcus thermophilus* bacteriophage DT1 (Tremblay & Moineau, 1999) (Accession number NC_002072). But in DT1, and also in other *Streptococcal* and *Lactococcal* phages, e.g. phages Sfi21 (NC_000872), SM1 (NC_004996) and BK5-T (NC_002796), this ORF is located upstream of the *cos* site. Orf2 through Orf15 constitute the DNA packaging and morphogenesis module. In general, the ORFs from this module showed high similarity to those of phages from related organisms, e.g. *Streptococci*, *Lactococci* and *Staphylococci*. Most of the structural proteins showed similarity to structural proteins from *S. thermophilus* phage DT1 (Fig. 2)

## Lysis module

The packaging and morphogenesis module is followed by the lysis module. Orf18 showed weak similarity to a putative holin from *Streptococcus suis* (Table 1). Orf19 contained two glycohydrolase domains, which indicates that this protein could act as endolysin and cleave the glycosidic *N*-acetylmuramoyl-(β1,4)-*N*-acetylglucosamine bond of the sugar backbone of peptidoglycan. This is supported by the similarity over the first *c.* 200 amino acids to muramidases encoded by bacteriophages from low GC gram-positive organisms (Fig. 3), including the lytic enzyme from *Streptococcus pneumoniae* bacteriophage Cp-1, whose structure has been solved (Hermoso *et al.*, 2003). The acidic residues of Cpl-1 thought to function in catalysis were conserved in Orf19. In general, the C-terminal domains of endolysins specify binding to the cell wall (Loessner, 2005). The C-terminal part of Orf19 showed no similarity to other proteins, except for the C-terminal part of a putative endolysin from *Streptococcus pyogenes* MGAS10394 (accession number YP_060444).

The stop codon of Orf19 overlapped with the startcodon of Orf20, which was similar to putative endolysins that appear to be encoded predominantly in bacteriophages from pyogenes *Streptococci* and *Staphylococci* (Fig. 4). Orf20 contains a CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) domain, which is present in a variety of

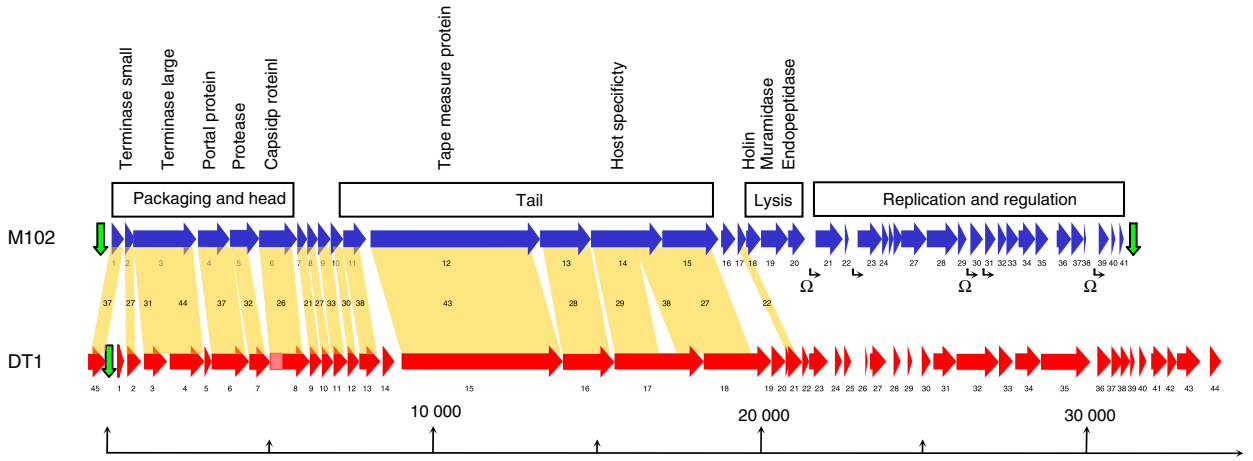


**Fig. 2.** Organization of *Streptococcus mutans* phage M102 genome and comparison with the genome from *Streptococcus thermophilus* phage DT1 (Accession number NC_002072). For clarity, Orf45 from DT1 is shown at the left site. Putative promoters are indicated with horizontal arrows, *cos* sites by vertical arrows. Putative terminators are depicted by ΩOrfs with similarity are connected by shading and the similarity (in % amino acid identity) is indicated.

**Table 1.** Characteristics of Orfs encoded by bacteriophage M102

| ORF | Start | Start codon | Stop | Size (aa) | pI | % identity (over aa) | e-value | With [organism] (Genbank accession number) | Putative function (conserved domain) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 181 | ATG | 549 | 122 | 9.61 | 36 (95) | 3e-12 | Restriction endonuclease [*Pediococcus pentosaceus* ATCC 25745] (ABJ68063) | Endonuclease (HNHc) |
| 2 | 571 | ATG | 909 | 112 | 5.29 | 27 (96) | 0.046 | Putative small subunit terminase [*S. thermophilus* bacteriophage Sfi19] (AAD44055) | Small subunit terminase |
| 3 | 896 | ATG | 2770 | 624 | 5.17 | 40 (614) | 1e-133 | ORF22 [*S. thermophilus* bacteriophage 7201] (AAF43515) | Large subunit terminase (terminase_1) |
| 4 | 2840 | ATG | 3784 | 314 | 5.44 | 37 (295) | 1e-48 | Putative portal protein [*S. thermophilus* bacteriophage DT1] (AAD21882) | PORTAL protein (Phage_portal) |
| 5 | 3781 | TTG | 4677 | 298 | 4.09 | 31 (195) | 6e-19 | Predicted Clp-protease [*S. thermophilus* bacteriophage Sfi21] (AAD41032) | CLP protease (CLP_protease) |
| 6 | 4697 | ATG | 5833 | 378 | 4.82 | 24 (368) | 2e-15 | Major structural protein [*Lactococcus* phage BK5-T] (AAK56807) | Phage capsid (phage capsid) |
| 7 | 5876 | GTG | 6187 | 103 | 4.15 | 33 (93) | 5e-04 | DNA packaging, phage associated [Bacteriophage Sal2] (YP_535183) | DNA packaging protein |
| 8 | 6184 | ATG | 6528 | 114 | 9.35 | 27 (62) | 0.06 | ORF28 [*S. thermophilus* bacteriophage 7201] (AAF43521) | Unknown |
| 9 | 6521 | ATG | 6913 | 130 | 6.10 | 32 (115) | 2e-09 | ORF29 [*S. thermophilus* bacteriophage 7201] (AAF43522) | Unknown |
| 10 | 6897 | ATG | 7253 | 118 | 4.30 | 30 (92) | 3e-05 | Probable tail component protein 123 [*S. thermophilus* phage Sfi19] (T09268) | Probable tail component |
| 11 | 7271 | ATG | 7954 | 227 | 5.52 | 38 (194) | 3e-34 | MPS-7201 [*S. thermophilus* bacteriophage 7201] (AAB71820) | Probable major tail protein (phage tail) |
| 12 | 8127 | ATG | 13 328 | 1733 | 10.13 | 43 (1754) | 0 | Putative tail component protein [*S. thermophilus* bacteriophage DT1] (AAD21891) | Tail protein (CHAP; pfam05257) (LT-GEWL; pfam1464) (Phage-related tail protein) |
| 13 | 13 328 | ATG | 14 854 | 508 | 5.55 | 28 (512) | 4e-59 | Putative tail component protein [*S. thermophilus* bacteriophage Sfi21] (AAC39282) | Probable tail component |
| 14 | 14 851 | ATG | 17 007 | 718 | 4.81 | 29 (570)<br>39 (215) | 4e-58<br>1e-34 | Host specificity protein [*S. thermophilus* bacteriophage DT1] (AAD21894) | Host specificity protein |
| 15 | 17 008 | ATG | 18 765 | 585 | 4.94 | 27 (577) | 8e-50 | Structural protein [*Streptococcus* phage 2972] (AAW27943) | Minor structural protein |
| 16 | 18 786 | ATG | 19 289 | 167 | 7.48 | – | – | – | Unknown |
| 17 | 19 306 | TTG | 19 578 | 90 | 9.43 | 27 (83) | 0.001 | Hypothetical protein EJ-1p68 [Bacteriophage EJ-1] (NP_945307) | Unknown |
| 18 | 19 575 | ATG | 20 042 | 155 | 7.60 | 30 (107) | 0.003 | Phage holin, LL-H [*S. suis* 89/1591] (EAP40817) | Holin |
| 19 | 20 058 | ATG | 20 879 | 273 | 4.65 | 59 (220) | 8e-70 | Peptidoglycan endolysin [*S. agalactiae* bacteriophage B30] (AAN28166) | Cell wall hydrolase (Glyco_25) |
| 20 | 20 879 | ATG | 21 394 | 171 | 5.90 | 45 (140) | 2e-20 | Phage-associated cell wall hydrolase [*S. pyogenes* MGAS10394] (AAT87679) | Cell wall hydrolase (CHAP) |
| 21 | 21 720 | ATG | 22 538 | 272 | 9.60 | 54 (266) | 8e-70 | Hypothetical protein PEPE_1019 [*Pediococcus pentosaceus* ATCC 25745] (YP_804517) | Unknown |
| 22 | 22 571 | ATG | 22 753 | 60 | 9.42 | – | – | – | Unknown |
| 23 | 22 973 | ATG | 23 731 | 252 | 9.17 | 49 (259) | 2e-63 | ORF5 [*S. thermophilus* bacteriophage 7201] (AAF26604) | DnaC homolog (DnaC; COG1484.1) |
| 24 | 23 731 | ATG | 23 934 | 67 | 10.43 | 55 (49) | 1e-07 | Hypothetical protein lmo2272 [*Listeria monocytogenes* EGD] (CAD00350) | Unknown |
| 25 | 23 931 | ATG | 24 101 | 56 | 9.81 | – | – | – | Unknown |
| 26 | 24 094 | ATG | 24 342 | 82 | 3.96 | 32 (83) | 5e-04 | Hypothetical protein [Temperate phage PhiNIH1.1] (NP_438123) | Unknown |

**Table 1.** Continued.

| ORF | Start | Start codon | Stop | Size (aa) | pI | % identity (over aa) | e-value | With [organism] (Genbank accession number) | Putative function (conserved domain) |
|---|---|---|---|---|---|---|---|---|---|
| 27 | 24 352 | ATG | 25 098 | 248 | 7.29 | 35 (265) | 7e-35 | Conserved hypothetical protein – phage associated [*S. pyogenes* M1 GAS] (NP_269147) | (RecT; pfam03837) |
| 28 | 25 113 | ATG | 26 042 | 309 | 5.71 | 45 (340) | 4e-79 | Phage protein [Bacteriophage 9429.2] (ABF32000) | (PolC; COG2176.1) |
| 29 | 26 043 | ATG | 26 339 | 98 | 10.32 | 39 (92) | 3e-08 | Hypothetical protein, phage-plasmid associated [*S. thermophilus* CNRZ1066] (AAV62371) | Unknown |
| 30 | 26 477 | ATG | 26 821 | 114 | 9.88 | 29 (120) | 4e-05 | gp30 [Bacteriophage A118] (NP_463491) | Unknown |
| 31 | 26 912 | ATG | 27 193 | 93 | 10.38 | – | – | – | Unknown |
| 32 | 27 308 | ATG | 27 535 | 75 | 10.25 | – | – | – | Unknown |
| 33 | 27 541 | ATG | 27 936 | 131 | 10.39 | – | – | – | Unknown |
| 34 | 27 940 | GTG | 28 419 | 159 | 4.74 | – | – | – | Unknown |
| 35 | 28 416 | ATG | 28 811 | 131 | 4.84 | 87 (110) | 1e-48 | Putative single-stranded DNA-binding protein [*S. mutans*] (AAN59480) | Single-stranded DNA binding protein (SSB; pfam00436) |
| 36 | 29 074 | ATG | 29 511 | 145 | 8.56 | 40 (156) | 1e-24 | ORF10 [*S. thermophilus* bacteriophage 7201] (NP_038311) | Unknown |
| 37 | 29 504 | ATG | 29 908 | 134 | 7.26 | 38 (75) | 5e-06 | gp178 [*S. thermophilus* bacteriophage Sfi11] (AAF63066) | Unknown |
| 38 | 29 908 | ATG | 30 009 | 33 | 10.54 | – | – | – | Unknown |
| 39 | 30 367 | ATG | 30 711 | 114 | 4.97 | – | – | – | Unknown |
| 40 | 30 720 | TTG | 30 911 | 63 | 3.92 | – | – | – | Unknown |
| 41 | 31 027 | GTG | 31 101 | 24 | 10.32 | – | – | – | Unknown |

peptidoglycan cleaving enzymes with L-muramoyl-L-alanine amidases or D-alanyl-glycyl endopeptidase activity (Bateman & Rawlings, 2003).

Phage M102 therefore might encode two endolysins of different substrate specificity, a muramidase and an amidase or endopeptidase. The presence of two separate endolysins is rather unusual, although in many phages different specificities are combined in one polypeptide. A further peculiarity is the presence of a possible N-terminal signal sequence in Orf20, as predicted by the program SIGNALP (http://www.cbs.dtu.dk/services/SignalP/) (Fig. 4). The genes encoding Orf19 and Orf20 could be expressed in *Esherichia coli*, but both proteins accumulated in the insoluble fraction, which precluded confirmation of their role in host cell lysis (results not shown).

## Replication module

The region downstream from the lysis module most probably constitutes the replication module. Five putative promoters with high similarity to the $-35$ and $-10$ regions of the *E. coli* $\sigma^{70}$ consensus promoter were identified (Fig. 2). The putative promoters were located in the intergenic regions of the module that encodes proteins required for DNA replication. The sequences of four of these promoters were highly conserved among each other. The DNA replication module also contained three putative Rho-independent termination signals. They were located downstream of ORFs 20, 29 and 38 (Fig. 2). About half of the ORFs from the replication module had no homologs in the database. The ORFs showed no sequence similarity to ORFs from DT1, but some were similar to other *Streptococcal* bacteriophage ORFs that are implicated in replication (Table 1).

## Proteome comparison

The deduced amino-acid sequences of the ORFs encoded by M102 were compared with deduced amino-acid sequences from a collection of *Streptococcal* and *Lactococcal* phages (Table 2). Over the complete proteome, bacteriophage M102 was most closely related to bacteriophages 7201, Sfi21, Sfi19 and DT1 from *S. thermophilus*, which indicates that M102 belongs to the Sfi21-like siphophage group (Rohwer & Edwards, 2002). However, the similarity was largely confined to the similarity between the structural proteins (ORFs 1–18). For the remaining ORFs, phage M102 was more similar to *S. pneumoniae* phages MM1 and Ej-1 and to *S. pyogenes* phage PhiNIH1.1.

## Determination of the *cos* site

The *cos* site of phage M102 was estimated by comparison of the restriction enzyme pattern of heat-treated and

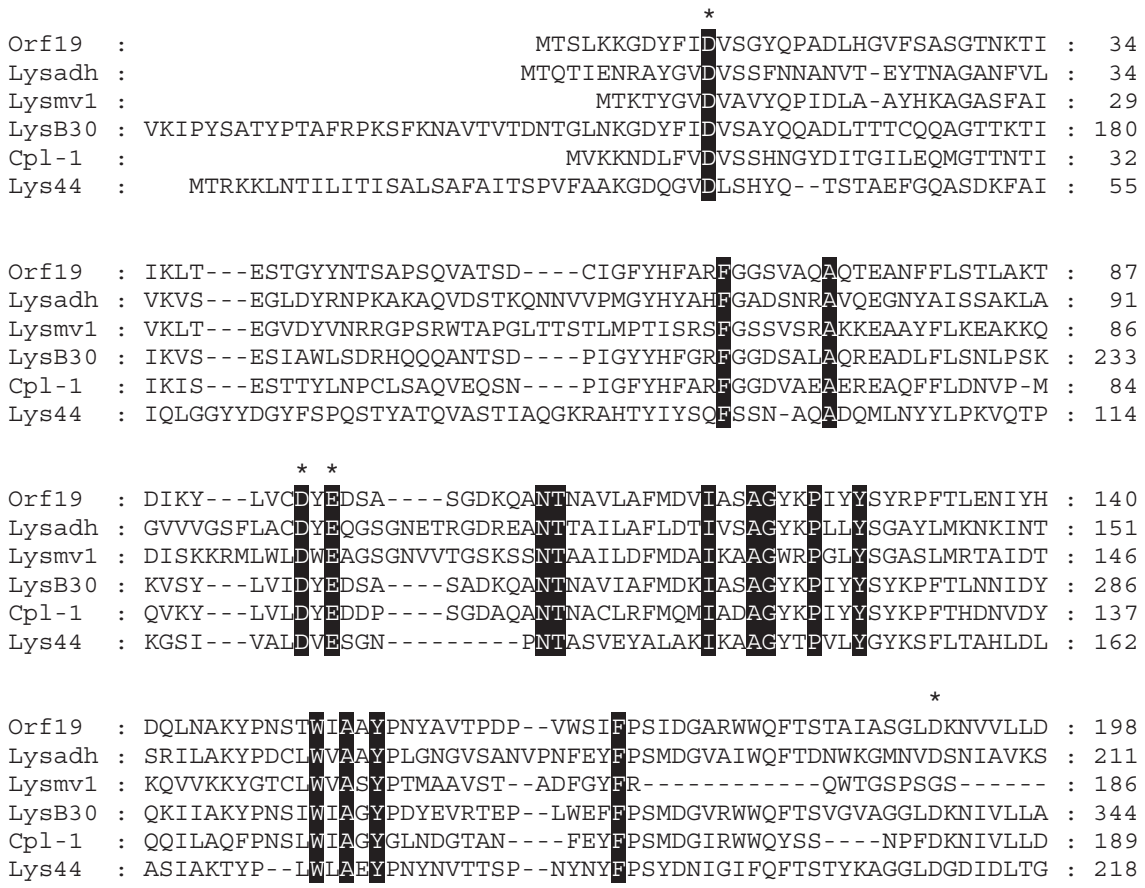


**Fig. 3.** CLUSTALW sequence alignment of Orf19 with bacteriophage lytic enzymes. The amino acid sequence of Orf19 from *Streptococcus mutans* phage M102 was compared with: Lys44, endolysin from *Oenococcus oeni* bacteriophage fOg44 (AAD10705.2) (Sao-Jose *et al*., 2000); Cpl-1, endolysin from *Streptococcus pneumoniae* phage Cp-1 (NP_044837.1) (Hermoso *et al*., 2003); LysB30, endolysin from *Streptococcus agalactiae* bacteriophage B30 (AAN28166.2) (Baker *et al*., 2006); Lysmv1, endolysin from *Lactobacillus bulgaricus* bacteriophage mv1 (P33486) (Boizet *et al*., 1990); Lysadh, lysin from *Lactobacillus acidophilus* bacteriophage ADH (NP_050170.1)(Henrich *et al*., 1995). Identical residues in all sequences are in white on a black background. The acidic residues of Cpl-1 proposed to function in catalysis (Hermoso *et al*., 2003) are indicated. Note that only partial sequences are shown.

nonheat-treated phage DNA. Digestion of nonheat-treated phage DNA with PstI gave one fragment of 6.6 kb, which resolved in two fragments of 1.4 and 5.2 kb upon heating. Digestion with XhoI gave a fragment of 4.0 kb, which became 0.7 kb smaller upon heating. Using a computer-generated restriction map of the sequence, the *cos* site could be mapped within a region of about 1 kb in size. For a more precise determination, heat-treated phage DNA and ligated phage DNA were directly sequenced using primers that were expected to hybridize closely to and in the direction of the expected *cos* site. Whereas the ligated phage DNA showed a contiguous sequence, the sequences of heat-treated phage DNA terminated, leaving a gap of 11 nucleotides (Fig. 5). The *cos* site of M102 thus has a 3′ overhang of 11 nucleotides (5′-ccgcgtgaata-3′). The stretch of the nine first nucleotides of this sequence is 89% identical to the last nine nucleotides from the *cos* site of *Streptococcus mitis* prophage SM1 (5′-gtgacggcgtgaa-3′) (Siboo *et al*., 2003).

## Comparison of bacteriophages M102 and M101

Bacteriophage M101 has the same host strain as M102, but the restriction pattern of M101 genomic DNA is different from that of M102 (results not shown). The lysis cassette and the *cos* site of M101 were amplified by PCR using primers derived from the M102 sequence sequenced and compared with M102. The *lysA* nucleotide sequence was 85.2% identical, whereas that of *lysB* was 82.6% identical. The LysA protein was 91.9% identical whereas the LysB protein was 91.8% identical. The *cos* sequence of M101 was identical to that of M102 (results not shown).


**Fig. 4.** CLUSTALW sequence alignment of Orf20 with bacteriophage lytic enzymes. Orf20 from *Streptococcus mutans* phage M102 was compared with: Skl, *N*-acetylmuramoyl-L-alanine amidase from *Streptococcus mitis* phage SK137 (CAJ13672) (Llull *et al*., 2006); PlyTW, *N*-acetylmuramoyl-L-alanine amidase from *Staphylococcus aureus* phage Twort (CAA69021) (Loessner *et al*., 1998); LysWMY, lysin from *Staphylococcus warneri* M phage ϕWMY (BAD83402) (Yokoi *et al*., 2005); Ply187, cell wall hydrolase Ply187 from *S. aureus* bacteriophage 187 CAA69022 (Loessner *et al*., 1999). Identical residues in all sequences are in white on a black background, conserved cysteine and histidine residues are indicated by an asterisk. The putative signal sequence in Orf20 is underlined. The complete sequence of Orf20 from M102 is shown, but partial sequences are shown from the other proteins.

**Table 2.** Proteomic comparison of M102 with *Streptococcal* and *Lactococcal* phages

| | | | Mean of protdist values* | | |
|---|---|---|---|---|---|
| Phage | Host | Accession number | 1–41 | 1–18 | 19–41 |
| 7201 | *S. thermophilus* | NC_002185 | 4.92 | 3.14 | 7.37 |
| Sfi21 | *S. thermophilus* | NC_000872 | 6.02 | 3.54 | 9.45 |
| Sif19 | *S. thermophilus* | NC_000871 | 6.05 | 4.05 | 8.82 |
| DT1 | *S. thermophilus* | NC_002072 | 6.24 | 3.52 | 10.00 |
| BK5-T | *L. lactis cremoris* | NC_002796 | 6.36 | 5.39 | 7.70 |
| MM1 | *S. pneumoniae* | NC_003050 | 6.90 | 7.25 | 6.41 |
| Lc-Nu | *L. lactis cremoris* | NC_007501 | 7.67 | 6.92 | 8.69 |
| EJ-1 | *S. pneumoniae* | NC_005294 | 8.41 | 9.17 | 7.37 |
| 315.1 | *S. pyogenes* | NC_004584 | 8.77 | 8.80 | 8.73 |
| O1205 | *S. thermophilus* | NC_004303 | 8.85 | 8.87 | 8.83 |
| PhiNIH1.1 | *S. pyogenes* | NC_003157 | 8.85 | 10.00 | 7.26 |
| Sfi11 | *S. thermophilus* | NC_002214 | 8.99 | 8.72 | 9.37 |
| 2972 | *S. thermophilus* | NC_007019 | 9.02 | 8.31 | 10.00 |
| Tuc2009 | *L. lactis cremoris* | NC_002703 | 9.15 | 10.00 | 7.97 |
| SM1 | *S. mitis* | NC_004996 | 9.45 | 9.55 | 9.32 |
| bIL170 | *L. lactis lactis* | NC_001909 | 9.47 | 9.09 | 10.00 |
| Cp-1 | *S. pneumoniae* | NC_001825 | 9.71 | 10.00 | 9.30 |

*The sum of all Protdist values was divided by the number of proteins used for the analysis. 1–41: all Orfs; 1–18: Orf1–Orf18; 19–41: Orf19–Orf41.

## Conclusions

In conclusion, the genome sequence of M102, which is the first from a bacteriophage that has *S. mutans* as host, shows high similarity to bacteriophages from *S. thermophilus* in the morphogenesis module, but less so or not in the lysis and replication modules. The lysis cassette of M102 is unusual in that it contains two lytic enzymes, one of which has probably an N-terminal signal sequence.

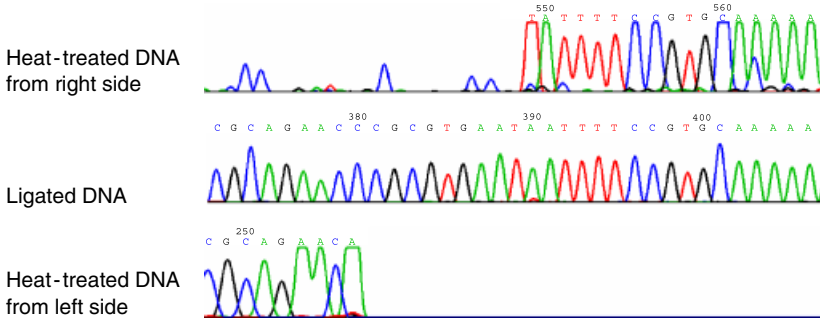


**Fig. 5.** Determination of the *cos* site of phage M102. The last base at each termination point of the sequence (T in the upper sequence and A in the lower sequence) is not present in the ligated DNA and caused by the DNA polymerase used for sequencing, which adds an A residue at the 3′ end.

## Acknowledgements

G. Tiraby is thanked for phages and host strain. The author appreciates the excellent technical assistance of Verena Osterwalder and Steve Reese. Bernard Guggenheim and Rudolf Gmür are acknowledged for support and critical reading of the manuscript, respectively. This work was supported by grant 215 from the Swiss Dental Association SSO.

## Nucleotide accession number

The nucleotide sequence reported here is deposited in the EMBL database under Accession number AM749121.

## References

Ajdic D, McShan WM, McLaughlin RE *et al.* (2002) Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc Natl Acad Sci USA* **99**: 14434–14439.

Armau E, Bousque JL, Boue D & Tiraby G (1988) Isolation of lytic bacteriophages for *Streptococcus mutans* and *Streptococcus sobrinus*. *J Dent Res* **67**: 121.

Bachrach G, Leizerovici-Zigmond M, Zlotkin A, Naor R & Steinberg D (2003) Bacteriophage isolation from human saliva. *Lett Appl Microbiol* **36**: 50–53.

Baker JR, Liu C, Dong S & Pritchard DG (2006) Endopeptidase and glycosidase activities of the bacteriophage B30 lysin. *Appl Environ Microbiol* **72**: 6825–6828.

Bateman A & Rawlings ND (2003) The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem Sci* **28**: 234–237.

Boizet B, Lahbib-Mansais Y, Dupont L, Ritzenthaler P & Mata M (1990) Cloning, expression and sequence analysis of an endolysin-encoding gene of *Lactobacillus bulgaricus* bacteriophage mv1. *Gene* **94**: 61–67.

Delisle AL & Rostkowski CA (1993) Lytic bacteriophages of *Streptococcus mutans*. *Curr Microbiol* **27**: 163–167.

Fischetti VA (2001) Phage antibacterials make a comeback. *Nat Biotechnol* **19**: 734–735.

Fischetti VA (2005) Bacteriophage lytic enzymes: novel anti-infectives. *Trends Microbiol* **13**: 491–496.

Greer SB, Hsiang W, Musil G & Zinner DD (1971) Viruses of cariogenic *Streptococci*. *J Dent Res* **50**: 1594–1604.

Henrich B, Binishofer B & Bläsi U (1995) Primary structure and functional analysis of the lysis genes of *Lactobacillus gasseri* bacteriophage phi adh. *J Bacteriol* **177**: 723–732.

Hermoso JA, Monterroso B, Albert A, Galan B, Ahrazem O, Garcia P, Martinez-Ripoll M, Garcia JL & Menendez M (2003) Structural basis for selective recognition of *Pneumococcal* cell wall by modular endolysin from phage Cp-1. *Structure* **11**: 1239–1249.

Higuchi M, Higuchi M & Katayose A (1982) Identification of PK 1 bacteriophage DNA in *Streptococcus mutans*. *J Dent Res* **61**: 439–441.

Hitch G, Pratten J & Taylor PW (2004) Isolation of bacteriophages from the oral cavity. *Lett Appl Microbiol* **39**: 215–219.

Klein JP & Frank RM (1973) Mise en évidence de virus dans les bactéries cariogènes de la plaque dentaire. *J Biol Buccale* **1**: 79–85.

Llull D, Lopez R & Garcia E (2006) Skl, a novel choline-binding *N*-acetylmuramoyl-L-alanine amidase of *Streptococcus mitis* SK137 containing a CHAP domain. *FEBS Lett* **580**: 1959–1964.

Loeffler JM, Nelson D & Fischetti VA (2001) Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science* **294**: 2170–2172.

Loesche WJ (1986) Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**: 353–380.

Loessner MJ (2005) Bacteriophage endolysins – current state of research and applications. *Curr Opin Microbiol* **8**: 480–487.

Loessner MJ, Gaeng S, Wendlinger G, Maier SK & Scherer S (1998) The two-component lysis system of *Staphylococcus aureus* bacteriophage Twort: a large TTG-start holin and an associated amidase endolysin. *FEMS Microbiol Lett* **162**: 265–274.

Loessner MJ, Gaeng S & Scherer S (1999) Evidence for a holin-like protein gene fully embedded out of frame in the endolysin

gene of *Staphylococcus aureus* bacteriophage 187. *J Bacteriol* **181**: 4452–4460.

Lukashin A & Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**: 1107–1115.

Rohwer F & Edwards R (2002) The phage proteomic tree: a genome-based taxonomy for phage. *J Bacteriol* **184**: 4529–4535.

Samaranayake LP (2002) *Microbiology of Dental Caries*, Churchill Livingstone, London.

Sao-Jose C, Parreira R, Vieira G & Santos MA (2000) The N-terminal region of the *Oenococcus oeni* bacteriophage fOg44 lysin behaves as a bona fide signal peptide in *Escherichia coli* and as a *cis*-inhibitory element, preventing lytic activity on oenococcal cells. *J Bacteriol* **182**: 5823–5831.

Siboo IR, Bensing BA & Sullam PM (2003) Genomic organization and molecular characterization of SM1, a temperate bacteriophage of *Streptococcus mitis*. *J Bacteriol* **185**: 6968–6975.

Tremblay DM & Moineau S (1999) Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* **255**: 63–76.

Tylenda CA, Calvert C, Kolenbrander PE & Tylenda A (1985) Isolation of *Actinomyces* bacteriophage from human dental plaque. *Infect Immun* **49**: 1–6.

Yokoi KJ, Kawahigashi N, Uchida M, Sugahara K, Shinohara M, Kawasaki K-I, Nakamura S, Taketo A & Kodaira K-I (2005) The two-component cell lysis genes *holWMY* and *lysWMY* of the *Staphylococcus warneri* M phage fWMY: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. *Gene* **351**: 97–108.